Graph Theoretical & Computational Problems in Structures & Rearrangement Reactions of Polycyclichydrocarbons (Graphs and Combinatorics III)

AUTHOR(S):

KAN, T.; TANAKA, N.; MIURA, H.; IIZUKA, T.

数理解析研究所講究録
397巻 1980年 16-29

16

Graph Theoretical & Computational Problems in Structures & Rearrangement Reactions of Polycyclichydrocarbons

Gakushuin Univ. T.Kan, N.Tanaka, H.Miura,
Gunma Univ. T.Iizuka

0. Introduction

(1) The standpoint of our study

The standpoint of our study is to abstract mathematical and/or computational problems from physical and/or chemical phenomena and/or problems, that is, to find some new concepts, theories and/or algorithms, and to solve the problems in question from the obtained view-point, or to give some predictive information to physisists and chemists.

(2) The aim of this paper

We shall introduce our graph theoretical works and pointout its computational or algorithmic aspects and essential problems in them.

(3) The problems in question

Extensive works on synthesis and rearrangement reaction of polycyclichydrocarbons have been done[1] but little knowledge seems to have been obtained about the structures of isomers. In fact, the structures of isomers of tricyclic hydrocarbon $C_{10}H_{16}$ have never been completely known until 1974. The composition of tricyclic hydrocarbon is generally expressed by C_nH_{2n-4}, and $C_{10}H_{16}$ is called adamantane. H.W.Whitlock et al.[2] studied rearrangement reactions among some isomers of $C_{10}H_{16}$ in 1968, but it is defective, because they failed to find all the isomers. A rearrangement reaction is a reaction between two isomers, that is, the composition is invariant under the reaction. The rearrangement reaction has been watched with interest as a new method of syntheses, since P.von Scheyer[1] succeeded in synthesizing adamantane.

However, it is well known that it is very difficult to synthesize polycyclichydrocarbons, because from the experimental point of view the structures of isomers are hardly known even in case that the number of cycles contained in an isomer is small(3,4or5). In fact, Sasaki group[3] and Hosoya group[4] studied some structures of chain, monocyclic and bicyclichydrocarbons. A.T.Balahan[5] studied the structures of isomers of $C_{2p}H_{2p}$ ($p \leqq 5$), but it seems that it is difficult to apply his method in case $p>5$. J.Lederberg[6] studied a system for computer construction, enumeration and notation of organic molecules as tree structures and graphs, but he did not completely classify the structures of isomers of hydrocarbon which has a given number k of cycles and a given number n of carbon atoms. Schleyer's group[1] is studying diamantane rearrangement reactions, which has the structures combining two structures of adamantane. Futhermore, McKervey[7] is studying triamantane rearrangement reaction, which has the more complicated structures combining three structures of adamantane.

However,these researchers are groping about in the dark, since they have not an algorithm finding all the structures of complicated molecules under some specified conditions.

On the other hand, the notation system for expressing polycyclic-hydrocarbons is becomming necessary, as such researches as mentioned above are increasing. For the present notation methods (CAS,Dyson,..,) fail to express complicated structures of polycyclichydrocarbons, namely more than one notations are given to a structure and the same notation is given to different structures.

(4) Graph Theoretical Problems

Considering the present status mentioned above, we think that it is significant to solve the following problems:

(a) to study algorithms finding the structures of isomers under some specified conditions such as a given number n of carbon atoms and a given number k of cycles,

(b) to abstract a graph theoretical concept from the chemical rearrangement reaction and to give some useful information about rearrangement paths to working chemists,

(c) to study notation or representation methods expressing the structures of polycyclic hydrocarbons.

We have adopted (a) as the first object, since it is the most fundamental problem and the essential or graph theoretical problem on (c) is automatically solved, if (a) is solved. Thus we shall describe our study about (a) and (b) in this paper.

Hereafter we shall deal with polycyclic hydrocarbons which have neather multiple bonds nor side chains, because we think that this case is subtential and the excluded cases are solved by operating some additional processes.

1. Structures of k-cyclic hydrocarbons

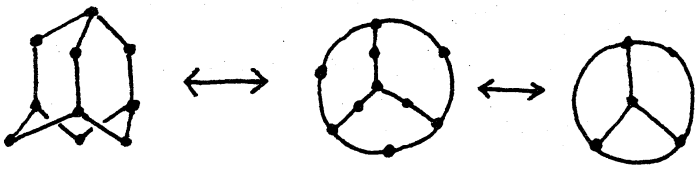
We are concerned about finding all the structures of isomers of k-cyclic hydrocarbons which have a given number n of carbon atoms, where "k-cyclic" means that the number of independent cycles in the isomer is equal to k. Our general principle is

(a) to find all the graphs or pseudographs which have neither degree 1 nor degree 2 vertices, we shall call such graphs or pseudographs as "frames" (see Fig.1), and

(b) to obtain all the structures of isomers by assigning degree 2 vertices to each edge of these graphs or pseudographs.

In fact, (a) is essentially a difficult graph theoretical problem, and if (a) is solved, (b) can be solved by using some computer method. In regard to (b), the efficiency of algorithm becomes a matter of great concern.

Fig. 1.



adamantane graph frame

1.1.Method I[8]

This consists of

(1) to classify frames having n vertices by using (d_3,d_4), where d_3 and d_4 are respectively the numbers of vertices of degree 3 and 4 ,

(2) to assign vertices of degree 2 to each edge of every obtained frame by using some inequality.

Now this method is outlined for tricyclichydrocarbons. Let G be a graph which has n vertices and m edges. We call G as a (n,m) graph. Remember the well known formula $k=m-n+1$, where k is the number of independent cycles in a (n,m) graph G. Since we deal with tricyclichydrocarbons, we put k=3. Thus we obtain $m=n+2$..... [1]

Let d_i be the number of vertices of degree i in G. Because we are concerned about hydrocarbons, i is equal to or less than 4. Then the degree sequence of G is (d_1, d_2, d_3, d_4). Therefore we obtain the following equations:

$$1 \cdot d_1 + 2 \cdot d_2 + 3 \cdot d_3 + 4 \cdot d_4 = 2m \quad \ldots\ldots [2]$$

and

$$d_1 + d_2 + d_3 + d_4 = n \quad \ldots\ldots [3] \ .$$

From [1] and excluding side chains, that is, $d_1 = 0$, we can drive the equations,

$$d_3 + 2 \cdot d_4 = 4 \quad \ldots\ldots [4]$$

and

$$d_2 + d_3 + d_4 = n \quad \ldots\ldots [5] \ .$$

Thus, for a given n we obtain the solutions (d_1, d_2, d_3, d_4) which satisfy both [4] and [5] (see Table 1.).

Table 1.

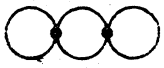
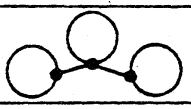
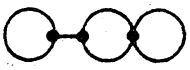
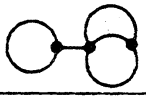
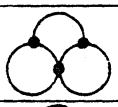
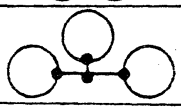
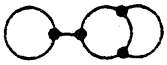
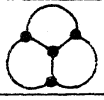
d_1	d_2	d_3	d_4
0	n-4	4	0
0	n-3	2	1
0	n-2	0	2

Then we introduce θ , the number of self-cycles in a frame, where a "self-cycle" means a cycle which has only one vertex of degree 3 or 4. Thus the three classes (d_1, d_2, d_3, d_4) in Table 1. are classified by θ into 2,4 and 4 minor classes respectively (see the θ column in Table 2.). Moreover we introduce Σ , the total number of vertices of degree 3 and/or 4 which belong to independent cycles in a frame. Then the 5th and 10th classes are divided by Σ into two minor classes respectively (see the Σ column in Table 2.). Thus we finally obtaine 12 frames as the solutions of [4] and [5] (see Table 2.).

Next we assign degree 2 vertices to each edge of these frames, namely a class of (n, n+2) graphs (see [1]) is constituted by assigning degree 2 vertices to each frame. In order to derive all the structures in

each class, we construct an index which expresses a manner of assigning vertices of degree 2. Consider the 1st frame in Table 2. as an example. We assign labels x_1, x_2, x_3 and x_4 to each edge, and for convenience, by x_1, x_2, x_3 and x_4 we also express the numbers of vertices on each

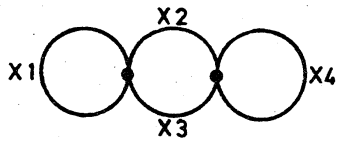
Table 2.

Hydrocarbon	Graph	d_1	(d_4, d_3)	θ	Σ	Class	No.
$C_n H_{2n-4}$	$(n, n+2)$	0	(2 , 0)	2	4		1
				0	6		2
			(1 , 2)	3	3		3
				2	4		4
				1	5		5
				0	6		6
				0	7		7
			(0 , 4)	3	3		8
				2	4		9
				1	6		10
				0	8		11
					9		12

Process for finding classes.

edge. We impose the conditions (see Fig.2.)

Fig. 2,



$$x_1 = x_4 \quad \text{and} \quad x_2 = x_3 \quad \ldots\ldots [6]$$

in order to avoid duplication by symmetry, and

$$x_4 = 2 \quad \text{and} \quad x_2 = 1 \quad \ldots\ldots [7] \ ,$$

since we exclude multiple bonds. Because the total number of vertices is equal to n and the frame in question is a solution (0,n-2,0,2) in Table 1. , we obtain

$$x_1 + x_2 + x_3' + x_4 = n - 2 \ \ldots\ldots [8] \ .$$

By finding all the nonnegative integral solutions (x_1,x_2,x_3,x_4) for the equations [6], [7] and [8], we can drive all the structures of isomers which belong to the class characterized by $(d_3,d_4,\theta\ ,\ \Sigma\ )= (0,2,2,4)$. We call a solution (x_1,x_2,x_3,x_4) as the "index" of the corresponding graph.

In a smilar way we can derive all the structures of isomers which belong to the class characterized by $(d_3,d_4,\ \theta\ ,\ \Sigma\ )=(4,0,0,9)$, and we show all the skeletons and indices in Table 3. , where a "skeleton" means a geometrical structure constructed by carbon atoms in a molecule.

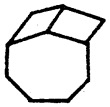
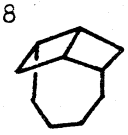
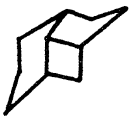
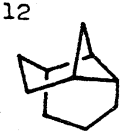
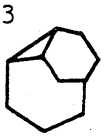
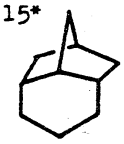
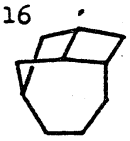
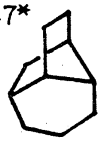
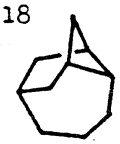
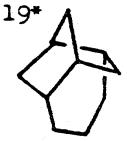
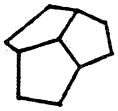
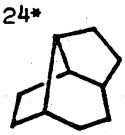
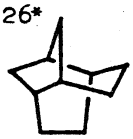
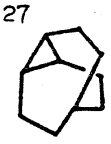
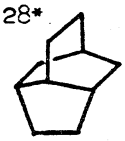
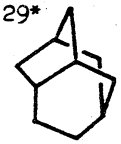
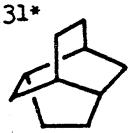
1·2 Method

This method is formulated by the expression

$$C(k) = B(k)+B(k-1)*B(1)+B(k-2)*B(1)*B(1)+\ldots\ldots+B(1)*B(1)\ldots\ldots*B(1) \ \ldots[9],$$

where C(k) is the set of k-cyclic frames and B(k) is the set of k-cyclic pseudoblocks. A "pseudoblock" is a frame which has no cut edge. And * means the connection operation by a cut point or a cut edge (see Fig 3).

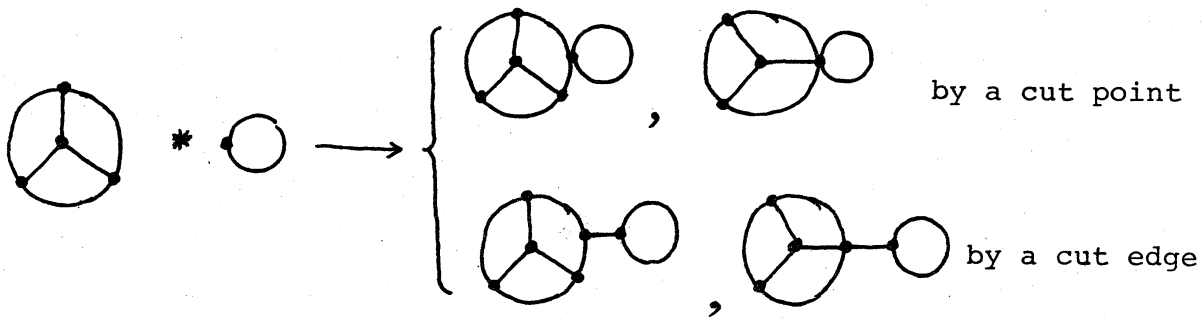
Table 3.

1  600000	2  510000	3  500100	4  420000	5  411000	6  410100	7  410010
8  410001	9  400200	10  330000	11  321000	12  320100	13  320010	14  320001
15*  311100	16  311010	17*  310200	18  310110	19*  310101	20  300300	21*  222000
22*  221100	23  221010	24*  220200	25  220110	26*  220101	27  220020	28*  211200
29*  211110	30*  211011	31*  210201	32*  111111			

* shows the isomers discussed in 2).

32* corresponds to Adamantane.

Fig. 3.



Namely, B(i)*B(j) is the set of all the frames which are generated by performing the operation* between each element of B(i) and each one of B(j). From the expression [9], it is clear that it is only necessary for obtaining C(k) to construct B(k), if B(1), B(2),....., B(k-1) are known. Now we show C(2) and C(3) in Fig 4 and Fig 5 respectively by using the expression [9]

Fig. 4.

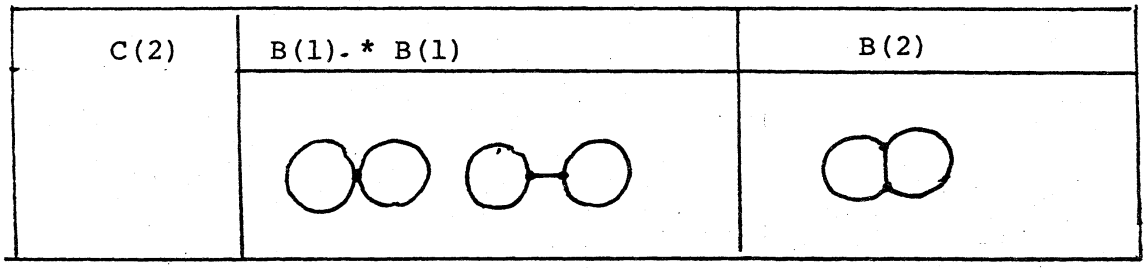
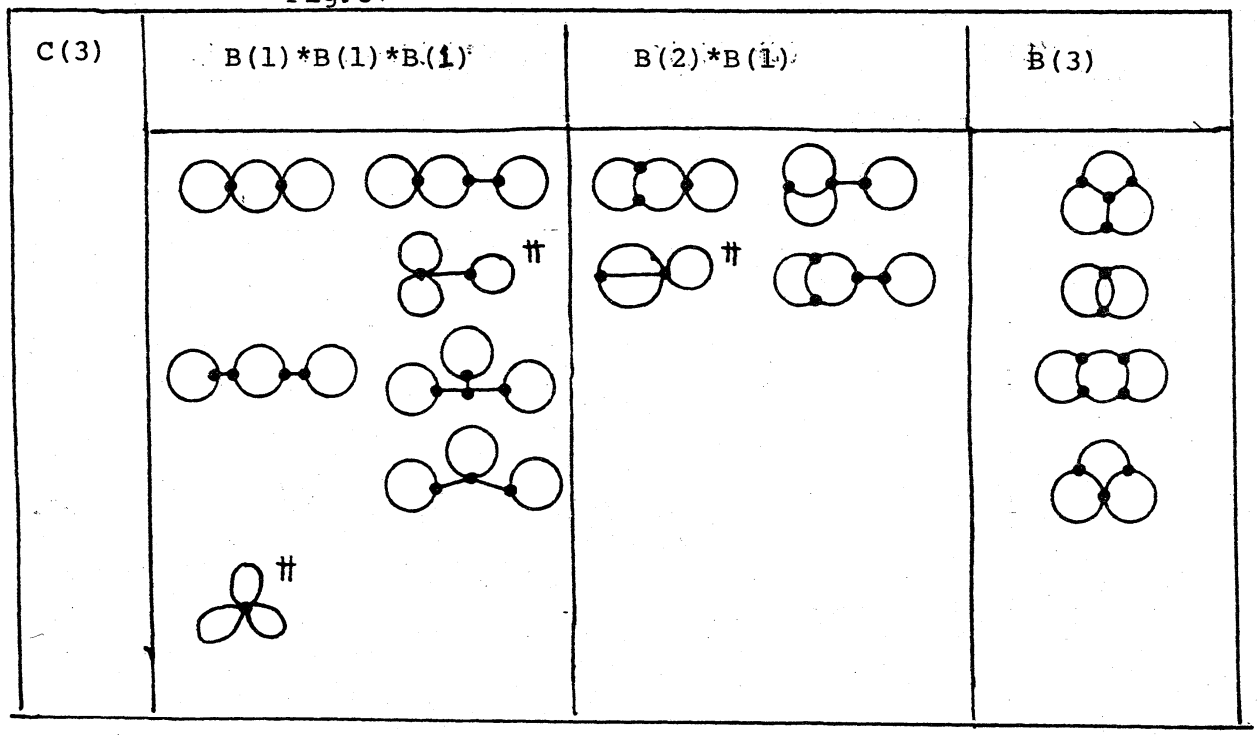


Fig. 5.



‡ : These frames are excluded in Table 2., since they contain vertices of more than 4 degree.

1.3 Algorithms and Results

(1) Algorithm by Method I

This algorithm was implemented for our early computer MELCOM 1101 and toward the end of 1973, first in the world, we obtained all the structures of isomers of 3-cyclichydrocarbons for n=10,11, where n is the number of carbon atoms. In fact, by this algorithm we can compute isomers for a given arbitrary n if necessary. However, this method can not be extended to more than 3-cyclic-hydrocarbons.

(2) Algorithm by Method II

This algorithm was implemented for our computer MELCOM 7500 in 1975 and we obtained all the structures of isomers of k-cyclic-hydrocarbons for k=4,5,6. In fact, by this algorithm we can compute isomers for a given arbitrary k.

(3) Algorithm by Method III

Method III is the latest, powerful and interesting method, but we don't describe it here, since the paper becomes too long. This algorithm was implemented for our present computer COSMO 700 in 1978, and we chequed all the previous results. Moreover, by using this algorithm we are investigating various cyclic structures without chemical restrictions.

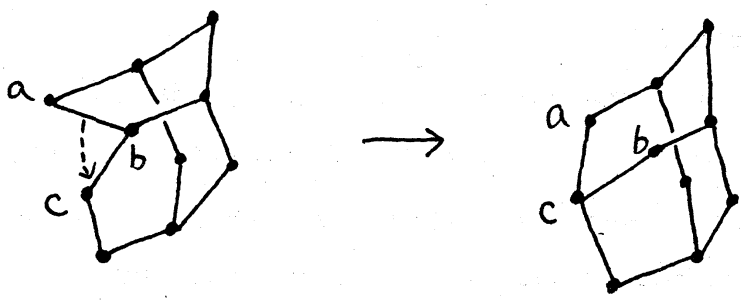
2. Rearrangement Reactions[9]

2.1 Graph Theoretical Formulation

(1) Transmutation

Remember the chemical phenomena:

an element t of T,



protoadamantane adamantane

this is a chemical rearrangement between isomers. Such rearrangements between isomers of 3-cyclichydrocarbons have been studied by Whitlock and Schleyer, since 1968. However, because it seems that they haven't an algorithm finding all the isomers of k-cyclic-hydrocarbons for even k=3, their studies have some defects. Now we formulate graph theoretically the rearrangement reaction as follows:

a transmutation T of a graph G is $T=\{t \mid G \xrightarrow{t} tG=(G-\{a,b\})+\{c,d\})\}$,

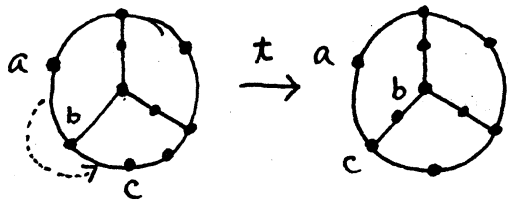
when t is an element of a transmuration T of G.

T is a restricted transmutation, if a=c.

T is a 1,n-transmutation, if T is restricted and dis(b,d)=n+1.

A 1,2-transmutation corresponds to a chemical 1,2-shift.

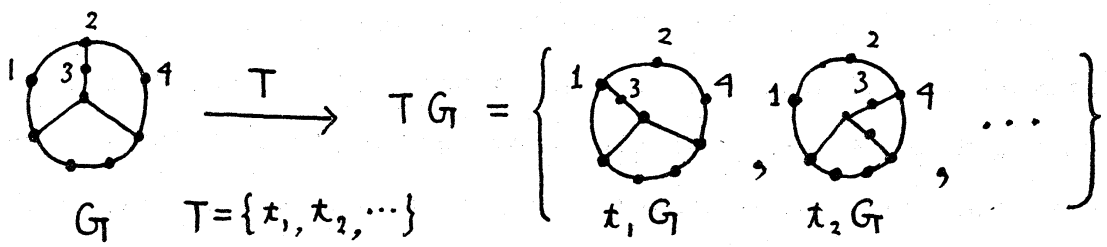
For example:



protoadamantane adamantane ,

and

the 1,2-transmutation T for protoadamantane ,



(2) Transmutation Paths

For given two graphs, a source G_s and a target G_t, a sequence $(G_0, G_1, \ldots, G_p)$ is a 1,n-transmutation path from G_s to G_t, if $G_0=G_s, G_p=G_t$ and $t_i G_i = G_{i+1}$ for $i=0,1,\ldots,p-1$, where t_i is an element of a 1,n-transmutation.

It is very important in chemistry to find a transmutation path, because it gives a prediction of an rearrangement path for a chemist.

2.2 Algorithms and Results

(1) Mono-source Propagation Algorithm (MSPA)

T is successively operated to a source G_s until $T^p G_s$ contains a target G_t. Thus a sequence $(G_s, t_1 G_s, t_2 t_1 G_s, \ldots, t_p t_{p-1} \ldots t_1 G_s = G_t)$ is a transmutation path. We obtained all the 1,2-transmutation paths between the isomers of adamantane (see Fig. 6.), we chequed the results by Whitlock and Schleyer, and could find their defects.

(3) Di-source Propagation Algorithm (DSPA)

T is successively operated to both G_s and G_t until $T^i G_s \cap T^j G_t = \emptyset$. Thus a sequence $(G_s, t_1 G_s, \ldots, t_i t_{i-1} \ldots t_1 G_s = t'_j t'_{j-1} \ldots t'_1 G_t, \ldots, t'_1 G_t, G_t)$ is a transmutation path. We could find some 1,2-transmutation paths between the isomers of diamantane (see Fig.7).

Comparing the number of graphs in each step of MSPA with that of DSPA, 1197 graphs in the case of MSPA and 118 graphs in the case of DSPA are generated to find all the transmutation paths between diamantane and No.6. The latter is one-tenth of the former! Furthermore, by SPA(Selective Propagation Algorithm)which doesn't be described here, we obtained a transmutation path for triamantane, perhaps first in the world. We informed this path as a prediction to Hamilton and Mckervey[7] who are chemically studing triamantane rearrangements.

3. Computational Aspects

(1) Efficiency

Efficiency is most important for algorithms. When devising an algorithm, we should consider a method by which computational quantity n!(the factorial of n) is reduced as small as possible, for example n^s (the s th power of n), where s is small natural number.

(2) Identification problem

This is to decide whether two graphs are isomorphic or not. This problem is concerned about all our algorithms. The speed of an algorithm depends on an adopted identification method.

(3) Symmetry

Efficiency of an algorithm can be improved by using symmetry of a graph. So it is very important to find symmetry of a graph.

(4) Internal Forms

It is necessary for improving efficiency of an algorithm to devise a suitable internal form of a graph.

Finally, we think, it is most essential to study a characterization method of vertices in a given graph for above mentioned problems.

Fig. 6

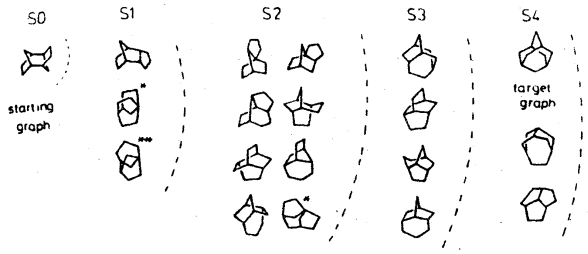
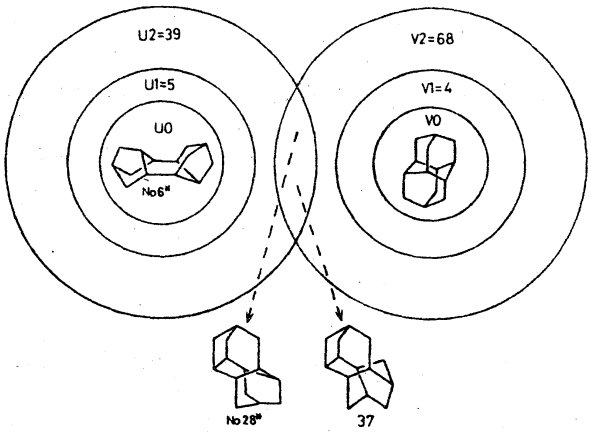


Fig. 7



References

1) R.C.Bingham and P.von Schleyer:"Recent Developments in the Chemistry of Adamantane and Related Polycyclic Hydrocarbons"; Topics in Current Chemistry 18, Springer-Verlag, 1972.

2) H.W.Whitlock and M.W.Sefken: J.Am.Chem.Soc., 98, 4929,1968.

3) Y.Kudo,: Kagaku no Ryoiki Zokan, 98, 115, 1972

4) K.Kawasaki, K.Mizutani and H.Hosoya: Natural Science Report Ochanomizu Univ., 22, 181, 1971.

5) A.T.Balaban: Rev.Roumaine Chim., 11, 1097, 1966.

6) J.Lederberg, Interim Report to NASA, 1966.

7) R.Hamilton, M.A.Mckervey, J.J.Rooney and J.F.Malone: J.C.S.Chem., 1027, 1976.

8) N.Tanaka, T.Iizuka and T.Kan: Chem. Lett.539, 1974.

9) N.Tanaka, T.Kan and T.Iizuka: J.of Chem. Inf.&Comput.Sci, 3, 162, 1979.